

ORIGINAL ARTICLE

Detection of mobile elements insertions for routine clinical diagnostics in targeted sequencing data

German Demidov¹  | Joohyun Park¹ | Sorin Armeanu-Ebinger¹ |
Cristiana Roggia¹ | Ulrike Faust¹ | Isabell Cordts² | Maria Blandfort³ |
Tobias B. Haack^{1,4} | Christopher Schroeder¹ | Stephan Ossowski¹

¹Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany

²Department of Neurology, Klinikum Rechts der Isar, Technical University Munich, Munich, Germany

³Neuropediatrics and Medical Genetics, Landau, Germany

⁴Centre for Rare Diseases, University of Tübingen, Tübingen, Germany

Correspondence

German Demidov, Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany.

Email: german.demidov@med.uni-tuebingen.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: 418081722 and 433158657; European Union's Horizon 2020 Research and Innovation Programme, Grant/Award Number: 779257

Abstract

Background: Targeted sequencing approaches such as gene panel or exome sequencing have become standard of care for the diagnosis of rare and common genetic disease. The detection and interpretation of point mutations, small insertions and deletions, and even exon-level copy number variants are well established in clinical genetic testing. Other types of genetic variation such as mobile elements insertions (MEIs) are technically difficult to detect. In addition, their downstream clinical interpretation is more complex compared to point mutations due to a larger genomic footprint that can not only predict a clear loss of protein function but might disturb gene regulation and splicing even when located within the non-coding regions. As a consequence, the contribution of MEIs to disease and tumor development remains largely unexplored in routine diagnostics.

Methods: In this study, we investigated the occurrence of MEIs in 7,693 exome datasets from individuals with rare diseases and healthy relatives as well as 788 cancer patients analyzed by panel sequencing.

Results: We present several exemplary cases highlighting the diagnostic value of MEIs and propose a strategy for the detection, prioritization, and clinical interpretation of MEIs in routine clinical diagnostics.

Conclusion: In this paper, we state that detection and interpretation of MEIs in clinical practice in targeted NGS data can be performed relatively easy despite the fact that MEIs very rarely occur in coding parts of the human genome. Large scale reanalysis of MEIs in existing cohorts may solve otherwise unsolvable cases.

KEYWORDS

genetic diagnostics, mobile elements insertions, structural variants, targeted sequencing

1 | INTRODUCTION

Mobile elements are segments of genomic DNA that can transpose to new regions of the genome, using either a cut-and-paste or copy-paste mechanisms, the latter leading to increased copy numbers. While the vast majority of mobile elements in the human genome are inactive and not capable of further expansion, some transposon families, in particular ALU, L1 and SVA, retained the ability to create novel insertions (Mills et al., 2007), i.e. mobile element insertions (MEIs). It has been estimated that ~1 out of 12–14 live births carries a *de novo* MEI genome-wide (Gardner et al., 2019). MEIs that affect exons and splice sites may impact the function of proteins, can cause genetic disease (Kazazian & John, 2017), and are detectable in whole exome sequencing (WES) data (Gardner et al., 2019; Torene et al., 2020). Moreover, deep intronic and regulatory MEIs can alter splicing and gene expression; however, their reliable detection requires whole genome sequencing (WGS; Kim et al., 2019), which is not yet widely used for clinical diagnostics.

In previous studies on rare diseases (RD), a moderate contribution of MEIs has been shown. For example, in 38,871 affected cases analyzed by WES (including 21,806 parent-child trios; Torene et al., 2020) MEIs were causal in 0.03% of all cases (95% CI: 0.02%–0.06%) and 0.15% of the cases with established molecular diagnoses (95% CI: 0.08%–0.25%). Another study of 9,738 WES parent-child trios (Gardner et al., 2019) revealed four likely causal *de novo* MEIs in 0.04% of cases (95% CI: 0.01%–0.11%).

Although the percentage of cases solved by MEI analysis is small, cohorts of several thousand RD patients are expected to contain one or more causal MEIs (Gardner et al., 2019; Torene et al., 2020). Hence, we see a need for re-analysis of existing RD cohorts and the implementation of MEI detection as a part of routine diagnostic pipelines in order to maximize diagnostic yield. Furthermore, guidelines for the computational evaluation of the functional consequences of MEIs are needed.

Mobile element insertions can also occur *de novo* during the lifetime of an individual (mosaic or somatic MEIs) and disrupt tumor suppressor genes in somatic cells, and MEIs have been shown to play a role in the development of cancer (Rodriguez-Martin et al., 2020). The current paradigm of personalized cancer medicine recommends the evaluation of somatic mutation patterns (e.g. mutational load, microsatellite instability, copy number variants) and cancer driver genes when selecting a patient's therapy. These include not only the more prominent gene fusions but also other translocation events such as MEIs.

In this study, we screened for rare MEIs in a cohort of almost 7.7 thousand individuals sequenced as part of routine clinical diagnostics, including 5,796 RD cases and

1,897 unaffected parents or siblings. Furthermore, we investigated 788 cancer cases, out of which 830 tumor and 788 paired-normal samples were sequenced by targeted gene panels for germline and somatic MEIs. We show that the number of candidate germline MEIs in RD cases is low after filtering for high quality and population frequency. We conclude that incorporation of MEI detection in a routine diagnostic analysis pipeline is feasible. A similarly manageable list of germline and somatic MEIs was detected per cancer case. We identified a causal MEI in one neurodevelopmental disease case and in one cancer predisposition syndrome case. Furthermore, we identified somatic MEIs affecting tumor suppressor genes in two tumor samples. In summary, germline MEIs caused ~0.03% of rare disease and cancer predisposition syndromes in our cohort, while somatic MEIs contributed driver mutations to 0.25% of cancer patients. This study is, to our knowledge, the first attempt to evaluate the potential impact of MEIs on cancer diagnostics in a routine clinical setting.

2 | METHODS

2.1 | Rare disease cohort analyzed by WES

In this study, we analyzed datasets from 5,796 RD patients and 1,897 unaffected relatives investigated by WES for routine diagnostics between 2017 and 2020. For 833 RD patients a parent-child trio and for 231 cases an unaffected relative were sequenced. Our WES cohort comprised patients with diseases of the nervous system (~35% of the affected cases), mental, behavioral or neurodevelopmental disorders (~31), familial neoplasms (~18%), endocrine, nutritional or metabolic diseases (~5%) and others. Age distribution of patients is shown in Figure S3. We scanned for MEIs in 2,504 WES samples prepared with Agilent SureSelect Human All Exon v6 and 5,189 WES prepared with Agilent SureSelect Human All Exon v7 hybrid-capture kits. Samples were processed uniformly and sequenced to around 130x average read coverage, as described previously (Froukh et al., 2020).

2.2 | Targeted sequencing of tumor and normal pairs

830 tumor and 788 matched normal tissues (mainly blood) from 788 cancer patients referred to the Molecular Tumor Board of the University Clinics Tübingen (Bitzer et al., 2020) were analyzed for therapeutically relevant molecular alterations including MEIs in coding regions,

introns or the close vicinity of cancer driver genes. Samples were sequenced between 2017 and 2020 with different versions of a custom targeted cancer gene panel. The different versions were based on the Agilent SureSelect Target Enrichment System and comprised between 337 genes (oldest version) and 708 genes (most recent version). Matched tumor and normal sample pairs were sequenced to an average coverage of 500× and 200×, respectively. Normal tissues were analyzed for germline mutations in cancer predisposition genes, while tumor tissues were used for the analysis of somatic variants. The complete procedure has been described by Hilke et al. (2020).

2.3 | Bioinformatics analysis and annotation of the NGS data

Generated sequences were processed using the MegSAP analysis pipeline (<https://github.com/imgag/megSAP>). MegSAP performs quality control, read alignment, various alignment post-processing steps, variant detection with freebayes (Garrison & Marth, 2012; SNV, short indels) and Manta (Chen et al., 2016; structural variants), as well as comprehensive annotation of variants with VEP (McLaren et al., 2016). Detection of copy number changes in germline and tumor sequencing data as well as tumor purity estimations were computed using ClinCNV (Demidov & Ossowski, 2019). Diagnostic analysis was performed using the GSvar clinical decision support system (<https://github.com/imgag/ngs-bits>). We used the GRCh37 reference genome assembly for the analysis.

2.4 | Detection and quality filtering of germline and somatic MEI events

For detection of germline MEIs in WES we used Scramble (Torene et al., 2020) with default filter settings. For high-depth tumor-normal pairs we applied stricter filter settings such as a threshold of more than 90% sequence identity between soft-clipped read clusters and the consensus MEI sequence, in order to improve removal of detected structural variants other than MEIs. Given that our samples had high coverage we selected only MEI sites covered by at least ten reads. Candidate MEIs covered with less than ten reads were evaluated visually using the Integrative Genome Viewer (IGV; Robinson et al., 2011). In vast majority, these changes were observed in off-target regions (far from vendor-specified enriched target regions) or likely represented artifacts that typically appeared as PCR-duplicate clusters or noisy soft-clipped reads without

a proper consensus sequence. MEIs occurring in close neighborhood (less than ten base pairs distance between start sites) in different samples were considered as a single MEI event.

Due to the differences in covered regions, samples were analyzed separately for the two exome-enrichment kits used for the RD cohort. Only rare MEIs were prioritized and we assumed that causal MEIs occur only within one affected family. Having a maximum number of four affected cases per family, we initially applied a threshold of maximally four observations of the same MEI in the whole cohort and subsequently removed MEIs observed in more than one family. An evaluation of more frequent MEIs (up to 20 per cohort) was also performed, however, no significant associations were found.

Tumor-normal data was processed using Scramble separately and scanned for both germline and somatic MEIs. Since we assumed that somatic MEIs are tumor-specific and unique we retained only MEIs occurring in four or less samples within the 830 tumor and 788 germline analyses for further evaluation. MEIs detected in both tumor and normal samples were labeled as germline MEIs, while MEIs only detected in tumor samples were labeled as somatic MEIs. MEIs detected only in the normal samples were discarded as potential false positives.

2.5 | Functional annotation and phenotype-based filtering of candidate MEIs

Each of the MEIs occurring next to OMIM-described genes were additionally evaluated using various annotations such as observed/expected score from GnomAD (Karczewski et al., 2020) and genotype-phenotype correlations provided by the Human Phenotype Ontology (Köhler et al., 2021). Additionally, detected MEIs were evaluated visually in IGV to remove soft-clipped read clusters produced by reads that likely represent PCR duplicates. Soft-clipped read clusters that were visually noisy (i.e. the alignments of the soft-clipped reads substantially differ from each other in terms of sequence and start site) were also filtered out. We provide a script based on (Greene et al., 2017) that takes as input a set of HPO terms describing a patient, and a gene-associated HPO set based on OMIM annotation, and returns a measure of the case-to-gene phenotype similarity (https://github.com/GermanDemidov/gene_to_phenotype_association). Phenotype similarity scores can be used as simple filter, or integrated as additional feature in clinical decision support systems such as GSvar.

3 | RESULTS

We identified 308 ultra-rare (4 or less events in the whole cohort, i.e. MAF <0.05%) candidate Alu, L1 and SVA sites in 375 samples of the RD cohort (375 individuals), comprising 401 distinct MEIs. Of these MEI sites, 142 were overlapping or in the close vicinity of known disease-associated genes listed in OMIM with a phenotype key of 3 (Amberger et al., 2019). In most cases (~70%), the elements were found in non-coding or intronic regions of genes associated with phenotypes unrelated to the patient's clinical presentations. Furthermore, we detected 96 candidate MEIs at 79 sites in 65 samples within the cohort of 830 tumor and 788 matched normal samples sequenced with the custom cancer panels. Eight candidate MEIs occurred in both tumor and normal tissue samples, hence our candidate list comprised 8 germline events and 80 somatic events. Seven germline MEIs were present in samples of more than one patient and are likely common MEIs or their impact remained unclear. Two somatic MEIs were detected in different tumor samples from the same patient.

3.1 | Detection of causal germline variants

The core phenotype of two cases were attributed to newly discovered germline MEI in *DNMT3A* (OMIM # 602769) and *RBI* (OMIM # 614041), respectively. Both cases had negative molecular diagnostic test results prior to this study. The variant in *DNMT3A* was found in a patient of the RD cohort, while the *RBI* variant was identified in a tumor-normal pair of the cancer cohort.

3.1.1 | Patient 1

In the WES data of patient 1, we identified an ALU element that was inserted directly into exon five of *DNMT3A* (ENST00000264709.3 c.465_466insALU, p.?) predicting a heterozygous loss-of-function of *DNMT3A*. Pathogenic *DNMT3A* variants have been associated with autosomal dominant Tatton-Brown-Rahman syndrome (TBRS, OMIM: # 615879). Reverse phenotyping revealed that key features observed in >80% of TBRS patients (Tatton-Brown et al., 2018) were present in our patient including exceptional overgrowth and intellectual disability. In addition, patient 1 was diagnosed with autism and other behavioral abnormalities, which were described in 51% of patients with TBRS. In line with a disease-causal role and postulated *de novo* status of the variant it was not detected in exome data of either of the healthy parents (Figure 1). No other pathogenic or likely pathogenic candidate variants

were identified in genes that have been previously associated with the patient's clinical features. Fragile X analysis, array CGH and conventional karyotyping were normal [Supplementary materials: Case Report 1].

3.1.2 | Patient 2

The second causal germline MEI was detected in the cancer cohort. In both the tumor and the normal data of patient 2, who was diagnosed with bilateral retinoblastoma [Supplementary materials: Case Report 2], we detected an ALU element insertion (ENST00000267163.4 c.940-17_940-16insALU, p.?) in *RBI*. We observed 24 and 31 soft-clipped reads supporting the exact same breakpoint in tumor and normal tissues, respectively. According to OMIM and HPO term associations the phenotype described for pathogenic *RBI* variants matched perfectly to the phenotype described for our patient. However, the identified MEI was intronic and the insertion site is 18 base pairs upstream of exon 10. Thus, it was unclear if it disturbed the correct splicing of the *RBI* mRNA and could be causal for the disease. To further evaluate our finding, we performed targeted gene panel sequencing of the unaffected parents, and RNA sequencing on patient's whole blood. We found no evidence for the presence of an *RBI* ALU insertion in the parental DNAs (Figure 2). This observation is in line with a *de novo* status of this MEI.

To investigate the consequences of such *de novo* intronic MEI on the correct splicing of *RBI*, we performed RNA-seq of a whole blood sample from patient 2. We found a multiple junction reads between exons 9 and 11 (17 split reads, Figure 3) indicating skipping of exon 10 (r.940_1049del, p.(Val314Phefs*2)). The change was not observed in RNA-seq data from ten randomly selected RNA-seq experiments of controls without retinoblastoma from our internal database.

In the tumor sample of patient 2, we furthermore found a loss-of-heterozygosity of the whole chromosome 13, where the *RBI* gene is located (computational purity estimation: 77.5%), which made this MEI homozygous in tumor. No other somatic driver mutations (SNVs, indels, CNAs) were identified in this tumor-normal pair, sequenced with our latest diagnostic panel of 708 cancer-associated genes. This observation and identification of a second hit in this gene, combined with *de novo* origin of the intronic MEI, and its impact on RNA, proves the causality of this particular variant for the observed phenotype.

3.1.3 | Other pathogenic variants involving sequences similar to mobile elements

In addition, Scramble identified several structural variants (SV) in both the RD and the cancer cohort, which are

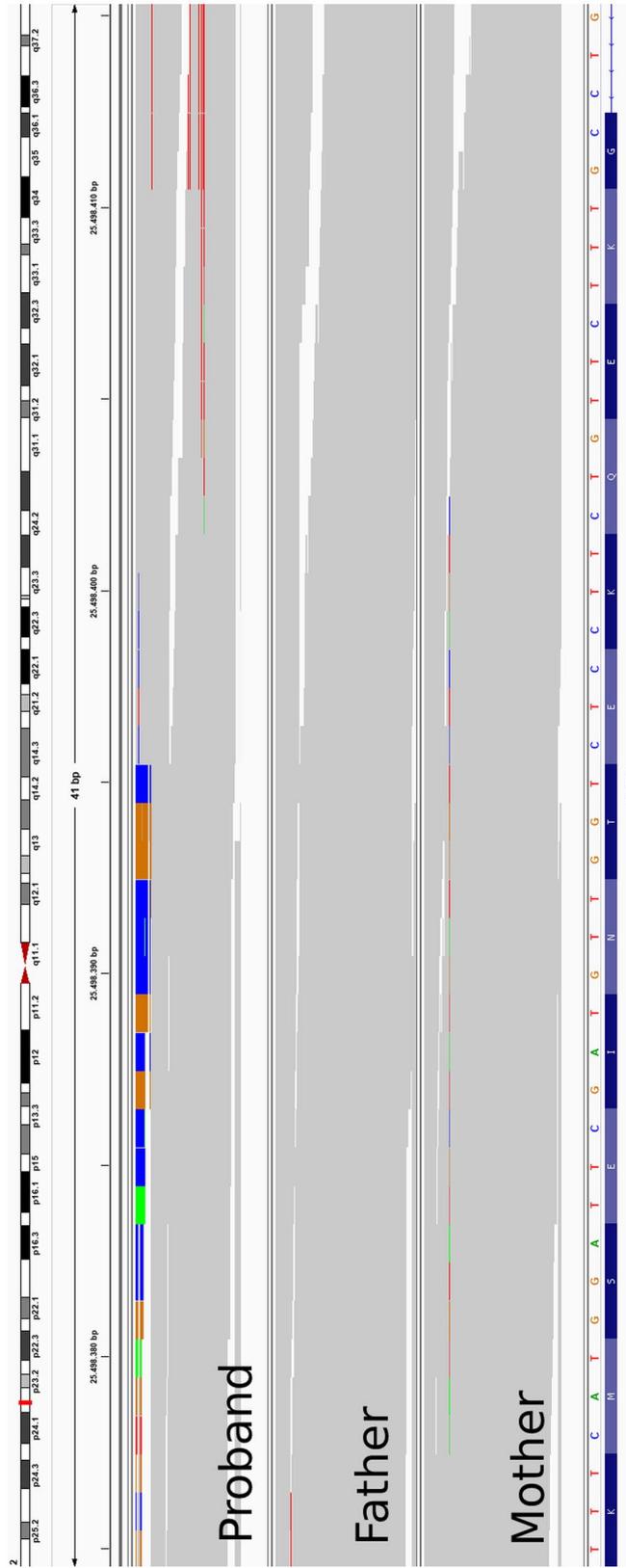


FIGURE 1 Read alignment (BAM) files of patient 1 (upper panel) and his parents (lower panels) in IGV browser. We observed a cluster of 10 soft-clipped reads with 100% sequence identity with ALU repeat sequence (start coordinate in MEI: 1, end in MEI: 46) showing the exact same breakpoint at the center of exon 5

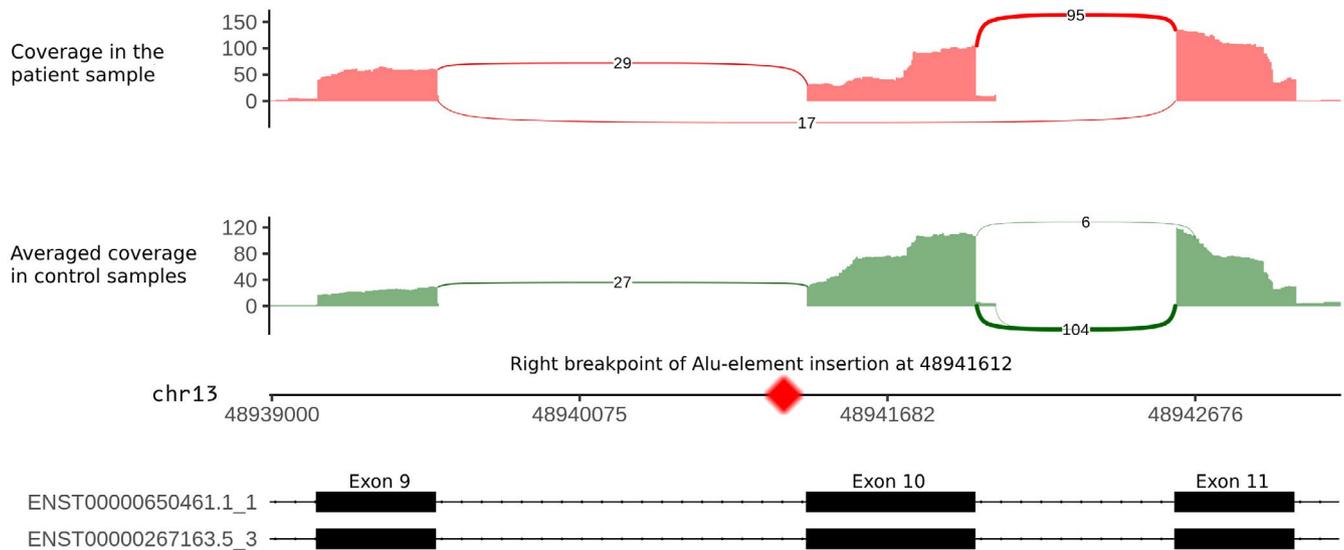


FIGURE 3 Sashimi plot (Garrido-Martin et al., 2018) of an exon skipping event caused by a *de novo* MEI variant in the intron of *RBI* upstream of exon 10. RNA-seq coverage data of patient 2 is shown in red in the upper plot, compared to the average coverage of 10 randomly selected control samples shown in green in the lower plot. Annotated exons of *RBI* in Ensemble are shown on the bottom

causal but not caused by a mobile element insertion. These events were picked up because the corresponding SV (e.g. a deletion) involved a sequence similar to a known mobile element sequence. For example, Scramble suggested a MEI insertion within exon 1 of *SPAST* (OMIM # 604277, ENST00000315285.9). The detected cluster of soft-clipped reads showed 94.3% similarity with an ALU-repeat, but the presence of stretched paired-end alignments and the configuration of soft-clipped alignments clearly indicated a deletion event instead of a MEI (coordinates in the GRCh37 reference genome: chr2:32,272,230–32,289,220). This deletion was also identified by CNV analysis using WES data, confirmed by MLPA analysis, and reported as causal variant to the clinician.

3.2 | Somatic MEIs in cancer driver genes

In the cohort of 788 cancer patients Scramble identified 96 high quality MEI candidates at 79 sites, of which 16 were found in both tumor and normal tissue and are therefore likely germline MEIs. By manual inspection of the remaining 80 somatic MEI candidates we identified and removed three MEIs that are germline but had a lack of coverage in the normal sample. Four MEIs were detected in a second biopsy of the same patient and are thus redundant. The remaining 73 somatic MEI at 73 unique sites were annotated using Cancer Genome Interpreter (Tamborero et al., 2018), labeling the events as deletions of the affected genes, since we expect a loss-of-function caused by the MEIs as the most plausible mechanism. CGI

indicates if the affected gene is a known driver of the respective cancer type and if a gain or loss of function is to be expected (tumor suppressor or oncogene). Subsequently, we excluded all MEIs that were predicted to be passengers or for which the affected gene and cancer type of the patient did not match with the known drivers reported by CGI. This filtering step resulted in 12 somatic MEIs that were further analyzed individually.

Out of the 12 candidates, five were filtered out due to QC issues (low quality of samples, variants supported by PCR duplicates only). One candidate was found to be a germline indel in *BARD1* (OMIM # 601593), which is frequent in the population (GnomAD 2.1 AF = 2.8%) and was accidentally picked up by Scramble due to sequence similarity of the affected region to mobile elements. Since the sequence similarity of this event was borderline to our threshold, it was filtered out in the normal tissue but kept in tumor. Another three clinically relevant candidates were likely driver structural variants, but different from MEIs, since the observed paired-end alignment signature was typical for deletion events.

Next, we investigated the remaining three somatic MEIs for their potential to drive tumor development based on the affected gene, the patient's tumor type and the CGI annotation. A partial insertion of LINE1 into the promoter of *CDKN2A* (OMIM # 600160) in a pancreatic tumor and a LINE1 insertion in exon 51 of *LRP1B* (OMIM # 608766) in a patient with esophageal carcinoma were considered as diagnostically relevant. An insertion of a LINE1 retrotransposon in intron 6 of the gene *WRN* (chr8:30924743) had unclear impact on the gene's function and was dismissed from further clinical interpretation.

The insertion in the promoter of *CDKN2A* (ENST00000579755.1, c.-47insL1) in a patient with neoplasm of the pancreas was considered as a likely loss of promoter function leading to the silencing of the tumor suppressor gene *CDKN2A*, which has been reported as a cancer driver and prognostic factor in pancreatic cancer (Doyle et al., 2019). Homozygous deletions (Lenkiewicz et al., 2020) as well as silencing by aberrant promoter methylation (Bernstein et al., 2015; Tang et al., 2015) of *CDKN2A* play a crucial role in pancreatic carcinogenesis. The insertion reported here is located close to the transcription start site in a highly regulated promoter region (likely disturbing binding of transcription factors). The insertion is accompanied by a loss-of-heterozygosity of this chromosomal region (coordinates of detected LOH in hg19 reference genome chr9:271606–28675984 as detected by ClinCNV [Demidov & Ossowski, 2019]), and is thus likely in a hemizygous state. We therefore classified the two-hit event as a homozygous loss of function of *CDKN2A*.

The low-density lipoprotein receptor-related protein 1B (*LRP1B*), which encodes the endocytic LDL-family receptor, is among the top 10 significantly mutated genes in human cancer and was associated with high tumor mutation burden (Chen et al., 2019). We considered the observed ENST00000389484.8: c.8202_8203insL1 in *LRP1B* as loss of function of this candidate tumor suppressor gene. LoF mutations in *LRP1B* have been frequently described in esophageal carcinoma (Sonoda et al., 2004).

4 | DISCUSSION

Here we have demonstrated the feasibility to detect and interpret MEIs based on clinical WES and targeted gene panel sequencing data generated in a routine diagnostic context. Concordant with previous studies, we also found that causal MEIs are extremely rare in RD but occur slightly more frequent as somatic driver events in tumor tissue. In 5,796 RD index cases we only identified one case clearly caused by a MEI inserted in *DNMT3A*. Interestingly, a second causal germline MEI was identified in the cancer cohort. Across the two cohorts we therefore found 2 out of 6,584 cases (0.03%, 95% CI: 0.004%–0.11%) caused by a germline MEI, consistent with previous estimations of 0.03% (Torene et al., 2020). In both cases the expected phenotype matched perfectly with the observed phenotype, showing that rigorous filtering and phenotype matching allows establish a diagnosis based on MEI events in a small but crucial portion of unsolved rare disease patients.

Notably, one of the causal germline MEIs in *RBI* was found in the cancer cohort by sequencing tumor-normal pairs with a 708-gene panel. In this case the availability of tumor data, in combination with the parental germline

DNA sequences and RNA-seq of the patient's blood allowed us to gather strong evidence for the causality of the MEI. Furthermore, it allowed us to investigate the functional consequences of the intronic MEI, i.e. an exon skipping, as well as to find a second hit (loss of chromosome) in the tumor genome. This suggests that paired tumor-normal analysis combined with RNA-seq could prove highly beneficial for the diagnosis of cancer predisposition syndromes in complex cases.

Most of the retinoblastoma-associated germline variants are nowadays efficiently found by sequencing of *RBI* (either by Sanger, targeted gene panel sequencing, WES, or WGS). However, a sizable fraction of cases remains unsolved (Lohmann & Gallie, 2018). Previous studies have also reported intronic MEIs as causal for *RBI* loss in familial retinoblastoma (Rodríguez-Martín et al., 2016). This suggests that *RBI* needs to be routinely analyzed for the presence of MEIs in patients with bilateral or positive family history retinoblastomas, but negative results in the standard *RBI* sequencing analysis.

Considering the genomic instability many cancers exhibit, clinically significant MEIs may occur more often in tumor tissues. However, with only 0.25% of cases we still found the percentage of cases with diagnostically relevant MEIs to be rather low. Although the number of candidate somatic MEIs was substantial (73 candidates), only two MEIs survived rigorous filtering and matching with known recurrently mutated driver genes in the respective cancer type. One additional case had unknown significance, while the remaining events were likely misinterpretations of other SV types such as deletions that affect a region with sequence similarity to mobile elements. Interestingly, all high quality MEIs found in somatic tissues were from the L1 family retrotransposons, in concordance with previous work (Rodríguez-Martín et al., 2020), further supporting that L1s are the most active mobile element in cancer.

Although structural variant callers such as Manta are in theory capable of detecting structural variants (SVs) in targeted sequencing data, MEIs that were found by Scramble were not detected by Manta, presumably due to the absence of the second corresponding paired-end cluster (second breakpoint) mapped to the reference genome. This highlights the need for using specialized tools for MEI detection, at least for targeted sequencing data analysis.

The computational runtime of Scramble for MEI detection in WES data is low. Moreover, using a large cohort for population allele frequency estimation, rigorous quality filtering and automated phenotype matching (OMIM, HPO) allows to efficiently prioritize MEI candidates in a semi-automated fashion without a significant increase in hands-on time for candidate evaluation. We therefore suggest that MEI analysis should be integrated into comprehensive routine diagnostic pipelines.

A main limitation of our study is the use of targeted sequencing, which has lower power in detection of diagnostically relevant novel insertions, especially in important non-coding parts of the genome such as introns or promoters. Indeed, despite the very limited coverage in these regions two out of four causal MEIs detected in our study reside in non-coding regions. Moreover, introns are generally much larger than exons, while mobile element insertions are large enough to disturb splicing even in deep-intronic regions. As suggested in previous studies (Gardner et al., 2019; Torene et al., 2020) we are likely underestimating the impact of MEIs in RD and cancer and additional studies, involving large cohorts of whole-genome sequenced samples, are required for evaluating the full potential of MEI detection in genetic disease diagnostics.

ACKNOWLEDGEMENTS

SO and GD received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257 (Solve-RD). T.B.H. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 418081722, 433158657. We acknowledge support by Open Access funding enabled and organized by Projekt DEAL. WOA Institution: N/A Blended DEAL : Projekt DEAL.

CONFLICT OF INTERESTS

The manuscript has been seen and approved by all the authors. The authors have no conflicts of interest or financial disclosures to report in relation with this work.

AUTHORS CONTRIBUTIONS

GD, SO: conceived and designed the study. GD performed bioinformatics analysis. JP, SAE, CR, UF, IC, MB, CR, CS, TH collected samples and clinical data, analyzed the clinical data and evaluated the clinical significance of variants. JP, SAE, CR prepared the case reports. GD and SO wrote the manuscript with the help of JP, TH, CS. SO, CS and TH supervised the study. All authors read and approved the manuscript.

ETHICAL COMPLIANCE

This study was approved by the decision of the Ethical Commission of Medical Faculty at Eberhard Karls University and University Clinics Tübingen (project number 066/2021BO2). Parents provided informed consent for publication with their affected children.

DATA AVAILABILITY STATEMENT

Variants details and scripts used to generate the analyses of this paper are available upon request unless that the type of request compromises ethical standards or legal requirements.

ORCID

German Demidov  <https://orcid.org/0000-0001-9075-4276>

REFERENCES

- Amberger, J. S., Bocchini, C. A., Scott, A. F., & Hamosh, A. (2019). OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Research*, *47*(D1), D1038–D1043. <https://doi.org/10.1093/nar/gky1151>
- Bernstein, D. L., le Lay, J. E., Ruano, E. G., & Kaestner, K. H. (2015). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *The Journal of Clinical Investigation*, *125*(5), 1998–2006. <https://doi.org/10.1172/JCI77321>
- Bitzer, M., Ostermann, L., Horger, M., Biskup, S., Schulze, M., Ruhm, K., Hilke, F., Öner, Ö., Nikolaou, K., Schroeder, C., Riess, O., Fend, F., Zips, D., Hinterleitner, M., Zender, L., Tabatabai, G., Beha, J., & Malek, N. P. (2020). Next-generation sequencing of advanced gliomas reveals individual treatment options. *JCO Precision Oncology*, *4*, 359. <https://doi.org/10.1200/PO.19.00359>
- Chen, H., Chong, W., Wu, Q., Yao, Y., Mao, M., & Wang, X. (2019). Association of LRP1B mutation with tumor mutation burden and outcomes in melanoma and non-small cell lung cancer patients treated with immune check-point blockades. *Frontiers in Immunology*, *10*, 1113. <https://doi.org/10.3389/fimmu.2019.01113>
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics (Oxford, England)*, *32*(8), 1220–1222. <https://doi.org/10.1093/bioinformatics/btv710>
- Demidov, G., & Ossowski, S. (2019). ClinCNV: Novel method for allele-specific somatic copy-number alterations detection. *BioRxiv*. <https://doi.org/10.1101/837971>
- Doyle, A., Kubler, M. M., Harris, A. C., López, A., Govindaraj, P., Prins, P., & Weinberg, B. A. (2019). The impact of CDKN2A mutations on overall survival in pancreatic adenocarcinoma. *Journal of Clinical Oncology*, *37*(4_suppl), 278–278. https://doi.org/10.1200/JCO.2019.37.4_suppl.278
- Froukh, T., Nafie, O., Al Hait, S. A. S., Laugwitz, L., Sommerfeld, J., Sturm, M., Baraghiti, A., Issa, T., Al-Nazer, A., Koch, P. A., Hanselmann, J., Kootz, B., Bauer, P., Al-Ameri, W., Abou Jamra, R., Alfrook, A. J., Hamadallah, M., Sofan, L., Riess, A., ... Buchert, R. (2020). Genetic basis of neurodevelopmental disorders in 103 Jordanian families. *Clinical Genetics*, *97*(4), 621–627. <https://doi.org/10.1111/cge.13720>
- Gardner, E. J., Prigmore, E., Gallone, G., Danecek, P., Samocha, K. E., Handsaker, J., Gerety, S. S., Ironfield, H., Short, P. J., Sifrim, A., Singh, T., Chandler, K. E., Clement, E., Lachlan, K. L., Prescott, K., Rosser, E., FitzPatrick, D. R., Firth, H. V., & Hurles, M. E. (2019). Contribution of retrotransposition to developmental disorders. *Nature Communications*, *10*(1), 4630. <https://doi.org/10.1038/s41467-019-12520-y>
- Garrido-Martin, D., Palumbo, E., Guigó, R., & Breschi, A. (2018). ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Computational Biology*, *14*(8), e1006360. <https://doi.org/10.1371/journal.pcbi.1006360>

- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. <http://arxiv.org/abs/1207.3907>
- Greene, D., Richardson, S., & Turro, E. (2017). ontologyX: A suite of R packages for working with ontological data. *Bioinformatics (Oxford, England)*, *33*(7), 1104–1106. <https://doi.org/10.1093/bioinformatics/btw763>
- Hilke, F. J., Sinnberg, T., Gschwind, A., Niessner, H., Demidov, G., Amaral, T., Ossowski, S., Bonzheim, I., Röcken, M., Riess, O., Garbe, C., Schroeder, C., & Forschner, A. (2020). Distinct mutation patterns reveal melanoma subtypes and influence immunotherapy response in advanced melanoma patients. *Cancers*, *12*(9), 12092359. <https://doi.org/10.3390/cancers12092359>
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. <https://doi.org/10.1038/s41586-020-2308-7>
- Kazazian, H. H., & Moran, J. V. (2017). Mobile DNA in health and disease. *The New England Journal of Medicine*, *377*(4), 361–370. <https://doi.org/10.1056/NEJMra1510092>
- Kim, J., Hu, C., Moufawad El Achkar, C., Black, L. E., Douville, J., Larson, A., Pendergast, M. K., Goldkind, S. F., Lee, E. A., Kuniholm, A., Soucy, A., Vaze, J., Belur, N. R., Fredriksen, K., Stojkowska, I., Tsytsykova, A., Armant, M., DiDonato, R. L., Choi, J., ... Yu, T. W. (2019). Patient-customized oligonucleotide therapy for a rare genetic disease. *The New England Journal of Medicine*, *381*(17), 1644–1652. <https://doi.org/10.1056/NEJMoa1813279>
- Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M., Callahan, T. J., Chute, C. G., Est, J. L., Galer, P. D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., ... Robinson, P. N. (2021). The human phenotype ontology in 2021. *Nucleic Acids Research*, *49*(D1), D1207–D1217. <https://doi.org/10.1093/nar/gkaa1043>
- Lenkiewicz, E., Malasi, S., Hogenson, T. L., Flores, L. F., Barham, W., Phillips, W. J., Roesler, A. S., Chambers, K. R., Rajbhandari, N., Hayashi, A., Antal, C. E., Downes, M., Grandgenett, P. M., Hollingsworth, M. A., Cridebring, D., Xiong, Y., Lee, J.-H., Ye, Z., Yan, H., ... Barrett, M. T. (2020). Genomic and epigenomic landscaping defines new therapeutic targets for adenocarcinoma of the pancreas. *Cancer Research*, *80*(20), 4324–4334. <https://doi.org/10.1158/0008-5472.CAN-20-0078>
- Lohmann, D., & Gallie, B. (2018). Retinoblastoma. GeneReviews® [Internet]. <https://www.ncbi.nlm.nih.gov/books/NBK1452/>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensemble variant effect predictor. *Genome Biology*, *17*(1), 122. <https://doi.org/10.1186/s13059-016-0974-4>
- Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends in Genetics*, *23*(4), 183–191. <https://doi.org/10.1016/j.tig.2007.02.006>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. <https://doi.org/10.1038/nbt.1754>
- Rodríguez-Martín, B., Alvarez, E. G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y. S., Temes, J., Garcia-Souto, D., Detering, H., Li, Y., Rodríguez-Castro, J., Dueso-Barroso, A., Bruzos, A. L., Dentre, S. C., Blanco, M. G., Contino, G., Ardeljan, D., ... Tubio, J. M. C. (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nature Genetics*, *52*(3), 306–319. <https://doi.org/10.1038/s41588-019-0562-0>
- Rodríguez-Martín, C., Cidre, F., Fernández-Teijeiro, A., Gómez-Mariano, G., de la Vega, L., Ramos, P., Zaballos, Á., Monzón, S., & Alonso, J. (2016). Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. *Journal of Human Genetics*, *61*(5), 463–466. <https://doi.org/10.1038/jhg.2015.173>
- Sonoda, I., Imoto, I., Inoue, J., Shibata, T., Shimada, Y., Chin, K., Imamura, M., Amagasa, T., Gray, J. W., Hirohashi, S., & Inazawa, J. (2004). Frequent silencing of low density lipoprotein receptor-related protein 1B (LRP1B) expression by genetic and epigenetic mechanisms in esophageal squamous cell carcinoma. *Cancer Research*, *64*(11), 3741–3747. <https://doi.org/10.1158/0008-5472.CAN-04-0172>
- Tamborero, D., Rubio-Perez, C., Deu-Pons, J., Schroeder, M. P., Vivancos, A., Rovira, A., Tusquets, I., Albanell, J., Rodon, J., Taberner, J., de Torres, C., Dienstmann, R., Gonzalez-Perez, A., & Lopez-Bigas, N. (2018). Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Medicine*, *10*(1), 25. <https://doi.org/10.1186/s13073-018-0531-8>
- Tang, B., Li, Y., Qi, G., Yuan, S., Wang, Z., Yu, S., Li, B., & He, S. (2015). Clinicopathological significance of CDKN2A promoter hypermethylation frequency with pancreatic cancer. *Scientific Reports*, *5*, 13563. <https://doi.org/10.1038/srep13563>
- Tatton-Brown, K., Zachariou, A., Loveday, C., Renwick, A., Mahamdallie, S., Aksglaede, L., Baralle, D., Barge-Schaapveld, D., Blyth, M., Bouma, M., Breckpot, J., Crabb, B., Dabir, T., Cormier-Daire, V., Fauth, C., Fisher, R., Gener, B., Goudie, D., Homfray, T., ... Rahman, N. (2018). The Tatton-Brown-Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants. *Wellcome Open Research*, *3*, 46. <https://doi.org/10.12688/wellcomeopenres.14430.1>
- Torene, R. I., Galens, K., Liu, S., Arvai, K., Borroto, C., Scuffins, J., Zhang, Z., Friedman, B., Sroka, H., Heeley, J., Beaver, E., Clarke, L., Neil, S., Walia, J., Hull, D., Juusola, J., & Retterer, K. (2020). Mobile element insertion detection in 89,874 clinical exomes. *Genetics in Medicine*, *22*(5), 974–978. <https://doi.org/10.1038/s41436-020-0749-x>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Demidov, G., Park, J., Armeanu-Ebinger, S., Roggia, C., Faust, U., Cordts, I., Blandfort, M., Haack, T. B., Schroeder, C., & Ossowski, S. (2021). Detection of mobile elements insertions for routine clinical diagnostics in targeted sequencing data. *Molecular Genetics & Genomic Medicine*, *9*, e1807. <https://doi.org/10.1002/mgg3.1807>